基于引用情感的潜在跨学科知识生长点识别研究

——以我国图书情报学科为例*

姜　霖[1]　张麒麟[2]
（1. 南通大学商学院，南通　226019；
2. 西南大学商贸学院，重庆　402460）

摘　要: [目的 / 意义] 随着科学研究问题的日趋复杂化和综合化，开展跨学科知识生长点的识别与研究，有助于把握当前学科发展现状和趋势，探索学科交叉演化规律。[方法 / 过程] 本文以我国图情领域 6 种学术期刊 2017—2021 年文献全文为研究对象展开实证分析。研究过程中利用 Bi-LSTM+CRF 模型对跨学科引用内容中的知识结合点进行识别，并尝试从引用情感角度对跨学科知识结合点的生长状况和趋势进行评估，从而为潜在跨学科知识生长点的发现提供帮助。[结果 / 结论] 研究结果表明，在图情领域研究中，跨学科概念、理论和工具的引入和结合与当前学科研究热点息息相关，跨学科方法的引入很大程度依赖于计算机科学中新算法模型的研究突破，并且其也开始对新数据环境下管理学方法的融入展开了探索。总体来看，图情学科知识与跨学科知识的结合点生命周期较短，缺乏长期稳定的跨学科知识引入点。

关键词： 跨学科　知识生长点　引文分析　引用情感
分类号： G202
DOI： 10.31193/SSAP.J.ISSN.2096-6695.2024.03.06

0 引言

“跨学科”概念最早由哥伦比亚大学伍德沃斯教授在 1926 年公开提出，是指超越一个已知学科的边界而进行的涉及两个或两个以上学科的实践活动[1]。现代学科在高度分化的基础上又呈现高度综合的趋势，伴随着学科间的相互渗透和彼此交融，学科间的界限逐渐被打破，学科边界也日益模糊。通过在复杂科学问题上运用综合性学科知识，往往能在原有的发展路径上生发出新

* 本文系国家社会科学基金青年项目“全文本引文分析视角下的交叉学科融合和演化机制研究”（项目编号：21CTQ040）的研究成果之一。
[作者简介] 姜霖（ORCID：0000-0003-3211-7783），男，讲师，研究方向为知识组织、知识挖掘，Email：jianglin@ntu.edu.cn；张麒麟（ORCID：0000-0001-9761-5463），男，副教授，研究方向为引文分析、知识组织，Email：zhang70sw@swu.edu.cn。

的生长点，成为新的科学前沿，也最有可能产生重大科学突破和革命性变革[2]。

Pierce[3]认为学科引用是科学文献之间传递学科信息的重要表现，反映了学科间的知识流动。如果引用内容中包含的某些跨学科理论或方法屡次为目标学科所引用，那么这些跨学科知识就很有可能与目标学科研究热点主题产生共鸣与协作，从而帮助目标学科消除发展过程中遇到的关键性瓶颈问题，新的跨学科知识生长点便由此产生，但对于哪些学科的哪些知识可以与目标学科的什么知识进行合作研究，仍然是科研人员面临的挑战。因此，为学科关键知识节点寻找跨学科相关知识，识别潜在的跨学科知识生长点，成为促进跨学科合作研究的根本[4]

1 相关研究

1.1 跨学科知识生长点概念研究

生长点的概念最初来源于植物学，本意是指具有强烈的细胞分裂能力的顶端分生组织[5]。李长玲等[6]借鉴自然科学中“科学生长点”的概念，将产生的创新知识点定义为“知识生长点”。他们认为知识生长点在学科原有知识外沿上呈凸起状，按知识来源可分为两类：(1)由本学科内部原有知识合作产生的新知识点，即学科内部知识生长点；(2)受跨学科知识的刺激，与本学科知识合作产生的新知识，即跨学科知识生长点。具体知识生长点产生原理如图1所示。

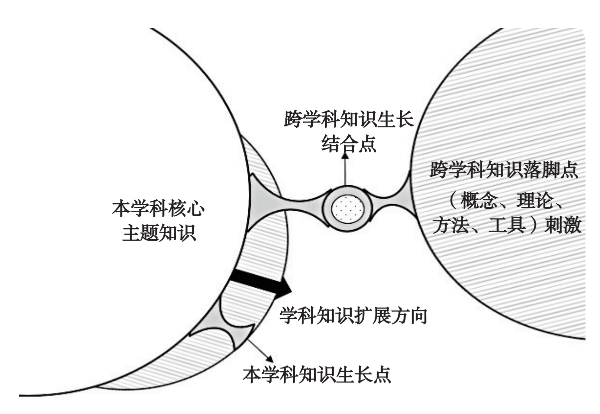


图1 学科知识生长点示意图

1.2 跨学科知识生长点研究

当前关于跨学科知识生长点的研究，大多通过引文数据和媒体数据的线索词、关键词、主题词等对跨学科知识流动和创新的动力、形式、机制展开研究。Porter等[7]指出，跨学科间知识的流动主要以学科之间概念、方法、理论等知识点的相互交换、转移、流动与扩散为基本特征。目前，多数学者通过识别目标学科知识与其他学科知识交叉的广度、强度、分布与扩散特征等[8]，来识别跨学科知识生长点受到的刺激程度，判断知识输出的作用和影响力以及知识点生长状态。如柯青等[9]利用直接引用率、布里渊指数以及逐年弹性系数三个指标，揭示了社会科学对图情领域研究的知识贡献推进效应及态势。Yegros等[10]对参考文献的丰富性、均衡性和差异性与知识输出影响力之间的关系进行了探讨。徐璐等[11]从输出强度、时效性、跨科学性三个维度对跨

学科引用中知识输出的效果进行了研究，发现跨学科引用有助于拓宽知识输出的学科范围，但不利于提高知识输出的实效性。周娜等[12-13]利用 LDA 主题模型和共现网络以传播学、图书情报学为例，对潜在的跨学科结合主题进行了探索。李长玲等[14-15]从跨学科引用率、多样性、集中度三个方面构建跨学科引用度模型，通过相关性分析对跨学科引用刺激知识生长的作用和程度进行了量化研究。荣国阳等[16]通过引文分析对不同主题下知识节点的影响力与共现强度进行了测量，并识别和分析了跨学科知识生长点所处的不同生命周期阶段。温芳芳等[17]选取 Web of Science 核心合集中以新冠肺炎为主题的论文为研究对象，利用关联规则方法挖掘新冠病毒领域的多学科知识融合特征，预测交叉知识点的生长方向。

当前跨学科知识生长点研究主要以预测跨学科合作主题为目标，但这类研究无法准确把握跨学科知识生长点的发展状态。部分研究从学科交叉测度出发，挖掘学科知识点受到的跨学科刺激程度，但学科交叉测度指标在探寻跨学科刺激及程度量化方面仍存在一定局限。

1.3 引用情感研究

引文内容中蕴藏着知识交流的具体内容，是研究知识流动的有效载体[18]。将引文内容引入分析框架，可以将不同学科间的知识流转过程具象化，为深入挖掘学科间的知识关联、知识吸收及内化过程提供条件。章成志等[19]通过对引文内容中蕴含的术语进行研究，验证了引用内容在学科交叉研究中的重要价值。

引用情感作为引用内容研究的一部分，表达了施引文献作者对被引文献的情感态度，可以进一步反映知识单元在学科知识体系中被接纳的程度。Moravcsik 等[20]通过对引文内容的人工解读来判断引文情感，并将其分为肯定引用、否定引用等五个维度，但受限于当时的技术水平，引用情感的识别效率低、主观性强。随着自然语言处理技术的发展，引用情感研究取得了长足进步。徐琳宏等[21]利用 SVM 算法识别引用内容中蕴含的引用情感并分为褒义、贬义和中性三类。Agarwal 等[22]提出了一种基于监督学习的引用情感自动分类框架来进行引用情感识别。廖君华等[23]利用 TF/IDF 值从引用内容中抽取情绪特征词汇并结合情感字典实现被引情感的识别。左任衔等[24]以 OpinionFinder 英文情感词典为基础，借助 SO-PMI 算法计算引文内容中词语的点互信息来构建引文领域情感词典。

综上所述，本文尝试从跨学科引用内容出发，通过术语识别技术探索引用内容中跨学科知识交叉的具体结合点，并利用情感量化分析模型更准确地把握潜在跨学科知识点的生长状况及动力因素。

2 跨学科知识生长点识别模型构建

2.1 研究思路

跨学科知识主要通过跨学科引用融入到目标学科知识体系中。本文首先抽取目标学科文献中包含的高频术语关键词作为学科知识的代表词汇，然后利用 BiLSTM+CRF 模型实现引用内容跨学科术语的识别，从而探寻目标学科知识与跨学科知识结合的落脚点，再借鉴图传播算法实现引用情感的量化，并以此判断跨学科知识结合点被接纳的程度以及能否成为新的学科生长点。具体实施路径如图 2 所示：

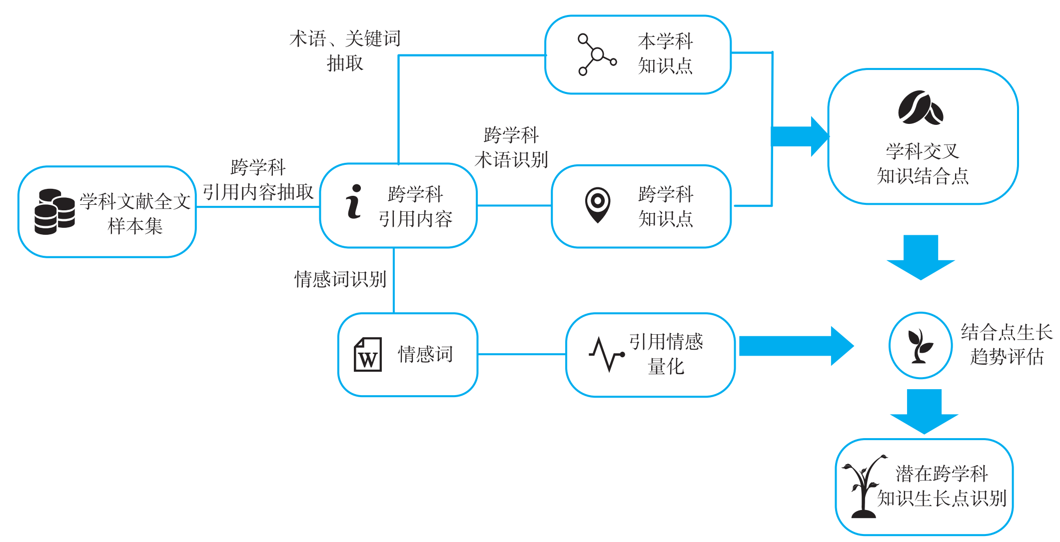


图 2　技术路线图

2.2　模型构建

（1）引用内容中跨学科知识结合点的识别

本文借鉴学科专业目录，利用人工标识参考文献来源期刊所属学科的方法识别跨学科引用。虽然在学科交叉融合的当下，学科界限日益模糊，但大部分期刊所刊载的学术论文依然集中在本学科领域，因此，通过期刊的学科归属来辨别跨学科知识来源具有一定的合理性和可行性，识别过程中默认综合性期刊属于目标学科。

领域术语是学科知识的核心表示。在学科知识体系中，概念常被认为是学科新思想或理念形成的先导因素，理论是推动研究范式形成和学科体系形成发展的决定因素，新研究方法和工具的引入，推动和加速了新知识体系的整合[25]。本文将引用内容中涉及的领域术语分为理论、概念（指标）、方法（模型）和工具（技术）四个类别。经实验比较，选择了 Transformer 大模型中的 BiLSTM+CRF 模型实现引用内容中跨学科知识术语的识别。

实验采用词标注法，使用 jieba 分词工具包和 BIO 三位标注法[26]实现语料数据的分词和标注。实验过程中选择的模型输入特征如下：

①词向量特征。词向量是一种词转换为稠密向量表示的技术，能够捕获词语之间的语义关系[27]。实验将文献数据集中的标题、关键词、摘要以及引用内容文本作为词向量模型的训练数据，并通过 word2vec 模型训练得到 300 维的词特征向量。

②词性特征。许多术语都遵循特定的构词规律，并通过其对应的词性表现出来。本文采用 jieba 分词工具中的词性标注体系对待识别文本内容中的词性进行标注。

③尾词和上下文特征。通过观察、归纳学科术语知识实体本身具有的语言学特征，将不同类型学科术语词汇具有的尾词特征和上下文特征整理如下：

尾词特征：“理论”“框架”“定律”“原理”“假设”“假说”“定律”“方法”“算法”“模型”“概念”“指标”“思想”“理念”“工具”“语料库”“软件”“平台”“数据库”等。

上下文特征：“提出”“基于”“根据”“采用”“利用”“结合”“引入”“借鉴”“运用”“借

助”等。

④学科领域特征。每个学科术语知识实体都具有其对应的学科属性特征。本文利用分词和词频统计方法寻找各学科引文内容中包含的高频关键词，并将之作为该学科的知识代表。按学科分类计算待识别词与各学科高频关键词之间的最大余弦相似度值，作为待识别词的多学科特征值（由于各学科包含的研究领域众多，不同领域间也存在较大差异，最大余弦相似值也就最能代表待识别词接近的学科领域），考虑到待识别词如“笔者”“学者”等普通类型词汇在各学科跨学科引用文本内容中分布会较为分散和平均，学科特征值标准差也会相对较低，易在识别过程中被过滤，因此将学科特征值的标准差作为综合学科特征属性值。

实验评价指标使用了准确率（Precision）、召回率（Recall）以及 F1 值（F-Measure），并采用五折交叉验证，模型中主要实验参数设置如表 1 所示。

表 1　实验参数设置表

参数名称及含义	参数值
epoch（迭代次数）	100
batch_size（批处理大小）	20
input layer（输入层）	324
hidden layer（隐藏层）	300
learning rate（学习率）	0.001
gradient clipping（梯度裁剪）	5
dropout（过拟合设置）	0.5
optimizer（优化器）	Adam 算法

实验识别效果如图 3 所示，方法术语识别效果最好，理论术语其次，其主要原因可能是方法术语和理论术语有较强的研究领域性特征，并且其语言学特征较为明显，例如“ERGM 模型”“时序低频词聚类算法”“齐普夫第二定律”等。概念术语识别准确率和召回率均一般，可能是由于概念术语较宽泛，语言学特征不明显且常存在多义性，如“熵”“半衰期”等既可以作为物理学概念，也可作为图情领域概念术语。工具术语识别效果较为不理想，主要是由于部分工具的适用性广且语言学特征也不明显，如“Wikidata Taxonomy Browser”“LIBSVM”等。

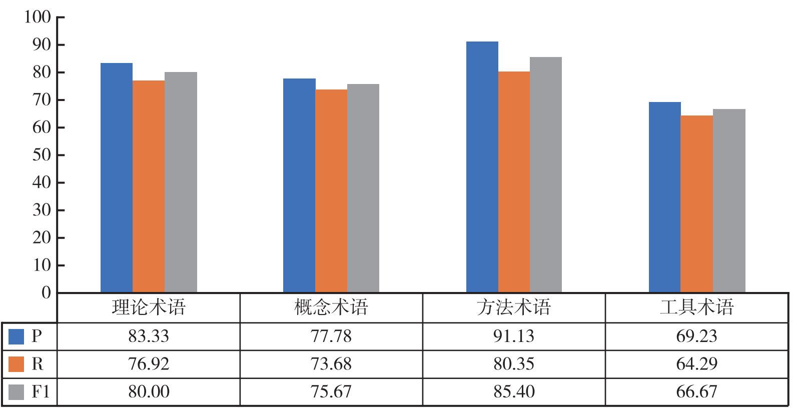


图 3　不同类型学科知识实体对比实验

本文利用深度学习算法实现引用内容中跨学科知识点的初步识别，并通过人工过滤方式保证跨学科知识词典的正确性。目标学科文献中高频关键词往往最能体现学科研究热点和本领域知识集合，因此本文将图情学科的高频关键词作为目标学科知识点代表。由于文献作者常在引用内容文本中选择使用代词指代学科知识点，因而当无法在引用内容中寻找到与跨学科知识相匹配的本学科高频知识点时，则以当前文献中的高频关键词作为本学科与跨学科知识的结合点。

（2）跨学科知识结合点生长趋势评估

本文在跨学科知识结合点生长趋势评估方面借鉴了前期研究成果[28]，通过图传播算法对引文内容中蕴含的引用情感进行识别和量化，从而将研究者对于跨学科知识点的情感评价转化为数值形式体现学科交叉形成过程中跨学科知识结合点的被接纳程度及生长状况。在图传播算法中，使用 word2vec 模型构建词向量，并利用词向量之间的余弦相似度值作为后续图传播算法模型中两词之间的图距离。在图传播模型中任何词与其自身的距离均设为 1。在情感词表自动构建初始，通过人工选择方式分别向正向、负向种子词集中添加该极性引用情感中表达程度最深的词。如可向正向种子词集中添加“大大提高”，向负向种子词集中添加“极差”，通过比较每个词到正、负向词集中词的平均图距离，计算词的情感极性值。详细算法步骤如下。

输入：假设输入为无向边加权图（G=V,E），其中 $W_{i,j} \in [0,1]$ 是边的权重，$(v_i, v_j) \in E$，V 表示包含在情感词典中的候选词集，G 表示节点之间的语义相似性。$P, N, \gamma \in \mathbb{R}$，$T \in \mathbb{N}$，P 代表正向情感种子集，N 代表负向情感种子集，γ 代表阈值，当词极性绝对值小于该阈值时，认为该词的词性为中性，T 代表候选词集中的词总数。

输出：　$pol \in \mathbb{R}^{|v|}$　　pol 代表情感极性

初始化：　$pol_i, pol_i^+, pol_i^- = 0$，对于所有的 i

$pol_t^+ = 1$，对于所有正向词集中的向量 $v_i \in P$

$pol_t^- = -1$，对于所有负向词集中的向量 $v_i \in N$

1. 设置 $\alpha_{ii} = 1$ 和 $\alpha_{ij} = 0 \quad i \neq j$

2. 对于向量 $v_i \in P$

3. $F = \{v_i\}$

4. 对于 t: 1...T

5. 对于 $(v_i, v_j) \in E$，所以 $v_k \in F$

6. $\alpha_{ij} = \max\left\{\alpha_{ij}, \alpha_{ik} \cdot \omega_{kj}\right\}$ $F = U\left\{v_j\right\}$

7. 对于所有的 $v_j \in V$

8. $pol_j^+ = \sum v_i \in P^{\alpha_{ij}}$

9. 重复 1–8，使用 N 来计算 pol^-

10. $\beta = \sum_i pol_i^+ / \sum_i pol_i^-$

11. $pol_i = pol_i^+ - \beta pol_i^-$

12. 如果 $|pol_i| < \gamma$ 就使 $pol_i = 0.0$

通过阈值的设定可以去除引用内容中包含的大部分中性词，较为准确地识别出情感词及强度。由于引用情感具有的复杂性和特殊性，如“训练时间长”“大量人工参与”等语义词组中，

单个词并不具有明显的情感倾向，“训练时间”“长”“大量”“人工参与”作为词组时才能体现引用情感的变化。综上，本文利用添加外部词表的方式，在分词时将这些特殊的词组视为一个整体识别其语义情感倾向和强度。

为了探寻引用情感指标的合理性，本文对采集的 6 种期刊的 5 年文献数据集中跨学科引用数量和引用情感系数进行了统计分析，具体统计结果和分布变化曲线如表 2 和图 4 所示。

表 2　跨学科引用年度分布情况表

指标名称	2017 年	2018 年	2019 年	2020 年	2021 年
跨学科引用量（次）	11 230	11 571	11 673	13 177	13 229
引用情感系数	1 862.046	1 844.814	1 946.938	2 129.113	2 235.009

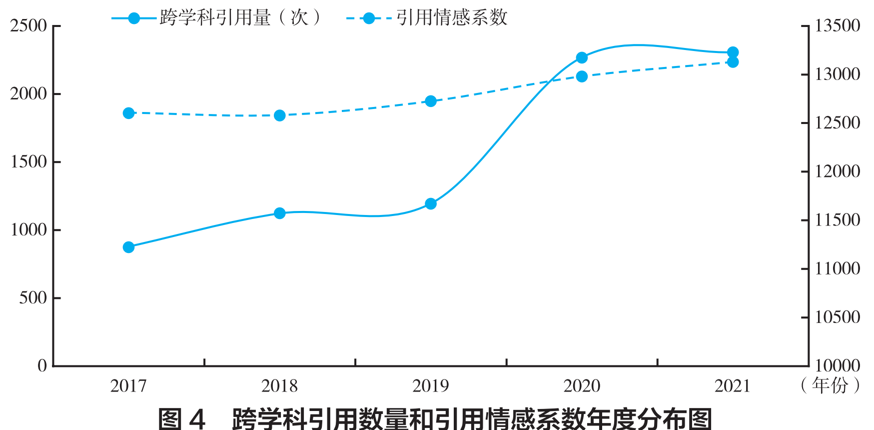


图 4　跨学科引用数量和引用情感系数年度分布图

利用 SPSS 统计分析软件对两项指标进行了斯皮尔曼秩相关分析，以反映指标间的相关程度[29-30]，结果如表 3 所示，在 0.05 水平（双侧）两项指标显著相关，相关系数为 0.9。此外，引用情感值离散系数为 0.077，相较于跨学科引用量指标更高，这也表明引用情感指标对于趋势变化的感知更显著，对跨学科知识在目标学科内部被接纳的程度感知更敏锐，能更好地反映跨学科知识结合点的生长状况。

表 3　斯皮尔曼秩相关分析表

指标名称	相关系数	方差	平均数	离散系数
跨学科引用量	0.9	851.441	12 176	0.070
引用情感系数		153.470	2 003.584	0.077

由于被引用行为本身也体现了对迁移知识点的认可，跨学科概念和工具在引用内容中较少涉及情感评价内容，所以本文赋予跨学科知识点基础情感评价系数 1，跨学科知识点的情感评价系数计算方法：$D_i = \sum_n F_i + \sum_n E_i$，其中 n 代表跨学科引用数量，F_i 代表跨学科知识点 i 出现的频次，E_i 代表跨学科知识点 i 的引用情感系数。具体的跨学科结合知识点生长状况评估方法示例如表 4 所示。

表4 跨学科知识结合点引用情感量化示例

参考文献	[26] 冯朝胜，秦志光，袁丁. 云数据安全存储技术[J]. 计算机学报，2015（1）:150-163.		
来源期刊	计算机学报	所属学科	工学（计算机科学）
引用内容	文章选择云平台对数据进行存储，但“云安全”缺乏一定的行业统一标准，云存储的安全性在学界和业界引起了广泛关注[26]，因此要重视平台数据的安全性		
知识结合点	信息安全 & 云安全；信息安全 & 云存储		
情感词	缺乏	引用情感量化	1-0.723 1
参考文献	[20] Cetin Y, Griffiths C, Ozel Z E, et al.Affective overload:the effect of emotive visual stimulion target vocabulary retrieval [J]. Journal of Psycholinguistic Research, 2016, 45（2）: 275-285.		
来源期刊	Journal of Psycholinguistic Research	所属学科	教育学（心理学）
引用内容	至此，情感负荷理论逐渐明晰，并陆续被应用到虚拟图书馆的信息需求[17]、信息检索行为[18]、信息获取的失败原因分析[19]、视觉刺激对用户认知的影响[20]等研究中		
知识结合点	信息需求 & 情感负荷理论；信息检索行为 & 情感负荷理论；信息获取 & 情感负荷理论；用户认知 & 情感负荷理论		
情感词	明晰	引用情感量化	1+0.343 2

3 实证分析

3.1 数据来源

作为学科专业性的刊物，学术期刊紧跟学术研究的前沿热点，发挥着引领学科发展的重要作用。本文以我国图情学科作为研究对象，选择了中文核心期刊目录（CSSCI）图情6种期刊作为具体数据来源，包括《中国图书馆学报》《情报学报》《图书情报工作》《情报资料工作》《情报杂志》《情报理论与实践》，选取知网（CNKI）作为学术文献资源获取平台，利用网络爬虫采集了这6种期刊2017到2021年发表的文献全文共6 971篇（已去除荐读、序、主编寄语等非规范文献）。

经识别统计，数据集中包含的规范性学术引用共132 461条，其中跨学科引用61 611条。采集到学科文献关键词共13 830个，选择其中频次大于5的关键词906个作为图情学科的代表关键词。

3.2 不同类型跨学科知识结合点增长分布

本文将跨学科迁移知识点分成了概念、理论、方法和工具四大类别，这主要是由于不同类别的跨学科知识在引用内容中出现的频次存在较大差异。从统计结果来看，我国图情领域研究中与跨学科知识的结合，更新速度较快，且会随着学科内部研究热点的变化而变化，但持续性跨学科知识输入较为匮乏。

为了探索跨学科知识结合的新兴增长点，表5统计了跨学科概念结合点年度增长分布情况，从中不难发现图情领域跨学科概念的引入来源较为广泛。2017年到2018年度，法学中数据治理、政府治理等概念的引入存在大幅增长。2018年以后，政府数据、科学数据的管理方式和方法成为了图情领域跨学科知识结合的热点方向。2019年到2020年度，随着新冠疫情的爆发，图情领域关注的热点也逐步向医疗健康领域转移，关于健康信息、公共卫生、信息规避等概念的引入大幅提升。

表 5　跨学科概念结合点年度增长分布表

跨学科概念结合增长点（top10）	引用情感系数		引用情感系数增加值
	2017 年	2018 年	
大数据治理 & 数据治理	2.000	54.145	52.145
政府决策 & 智库	3.000	37.863	34.863
安全信息 & 安全管理	4.000	30.905	26.905
网络空间 & 网络空间安全	0.000	20.977	20.977
高校图书馆 & 情绪	0.000	20.039	20.039
电子政务 & 公共服务	0.000	19.171	19.171
信息搜索 & 有限理性	1.000	19.960	18.960
影响因素 & 消费者	13.254	32.178	18.924
大数据治理 & 政府治理	0.000	18.493	18.493
人文社会科学 & 样本数据	0.000	18.225	18.225
跨学科概念结合增长点（top10）	引用情感系数		引用情感系数增加值
	2018 年	2019 年	
政府数据开放 & 数据开放	22.293	54.800	32.507
科学数据 & 数据开放	3.010	33.448	30.438
政府数据开放 & 政府数据	22.293	51.904	29.611
专利池 & 法律	1.000	24.527	23.527
移动图书馆 & 博弈	0.000	23.315	23.315
学术信息 & 情绪	0.000	22.450	22.450
知识库 & 本体	11.163	32.311	21.148
科学数据 & 元数据	2.189	22.869	20.680
术语抽取 & 依存关系	0.000	20.156	20.156
情报学 & 自我效能	0.000	20.040	20.040
跨学科概念结合增长点（top10）	引用情感系数		引用情感系数增加值
	2019 年	2020 年	
突发公共卫生事件 & 公共卫生事件	6.000	66.228	60.228
开放政府数据 & 政府数据	6.523	50.370	43.847
学科建设 & 国家安全	0.000	34.323	34.323
大数据治理 & 数据治理	9.485	42.999	33.514
在线健康信息 & 健康信息	1.408	33.795	32.387
专利价值 & 专利权	8.000	35.820	27.820
专利价值 & 专利质量	0.000	25.833	25.833
用户研究 & 信息规避	0.000	25.426	25.426
大数据治理 & 大数据	19.727	44.745	25.018
智能技术 & 人工智能	13.791	37.942	24.151
跨学科概念结合增长点（top10）	引用情感系数		引用情感系数增加值
	2020 年	2021 年	
科学数据 & 元数据	1.888	83.992	82.104
科学数据 & 数据开放	18.760	84.650	65.890
科学数据 & 数据资源	2.679	67.329	64.650
科学数据 & 生命周期	3.097	61.354	58.257
突发公共卫生事件 & 健康信息	0.000	26.969	26.969
健康信息行为 & 健康信息	9.394	36.127	26.733
影响因素 & 信息规避	0.000	26.399	26.399
健康信息行为 & 信息规避	1.000	25.988	24.988
技术创新 & 专利权	5.686	27.996	22.310
影响因素 & 情绪	1.000	22.881	21.881

表 6 统计了跨学科理论结合点年度增长分布状况，从中可以看到图情领域跨学科理论的引入主要来自管理学、经济学、心理学和传播学等领域。跨学科理论的引入和本学科研究主题之间存在较强的关联性。学科理论知识一般具有较高的学科属性，特定理论对于相关领域问题的解决具有重要作用。例如，图情学科知识贡献问题的研究主要借助于经济学理论，数据管理研究主要借鉴的是管理科学理论，信息行为研究主要依赖于管理学和心理学理论。

表 6　跨学科理论结合点年度增长分布表

跨学科理论结合增长点（top10）	引用情感系数		引用情感系数增加值
	2017 年	2018 年	
注意力机制 & 深度学习理论	0.000	13.019	13.019
注意力机制 &R 理论	0.000	13.019	13.019
注意力机制 &ACT-R 理论	0.000	13.019	13.019
知识贡献 & 社会认知理论	0.000	13.000	13.000
协同创新 & 行为理论	0.000	11.010	11.010
协同创新 & 系统动力学	0.000	11.010	11.010
知识贡献 & 社会交换理论	0.000	10.000	10.000
知识贡献 & 期望理论	0.000	9.000	9.000
知识贡献 & 社会资本理论	1.000	10.000	9.000
知识贡献 & 心理学理论	0.000	9.000	9.000
跨学科理论结合增长点（top10）	引用情感系数		引用情感系数增加值
	2018 年	2019 年	
情报学 & 心理学理论	10.858	29.666	18.808
知识传播 & 从众效应	0.000	11.635	11.635
专利价值 & 双边市场理论	0.000	11.545	11.545
情报学 & 社会分工理论	0.000	11.544	11.544
应用研究 & 整合理论	0.000	9.917	9.917
信息科学 & 心理学理论	0.000	9.114	9.114
信息系统 & 动机拥挤理论	0.000	8.928	8.928
政策变迁 & 多源流理论	0.000	8.382	8.382
政策变迁 & 政策变迁理论	0.000	8.261	8.261
演化分析 & 增量聚类框架	0.000	7.766	7.766
跨学科理论结合增长点（top10）	引用情感系数		引用情感系数增加值
	2019 年	2020 年	
使用数据 & 协同效应	0.000	14.470	14.470
突发公共卫生事件 &PMT 理论	0.000	12.000	12.000
游戏化 & 认知理论	0.000	10.654	10.654
人机交互 & 心理学理论	0.000	10.554	10.554
知识付费 & 行为理论	0.000	9.686	9.686
公共文化服务 & 文化资本理论	0.000	7.516	7.516
图书馆 & 学习理论	0.000	7.504	7.504
区块链 & 博弈理论	0.000	7.098	7.098
学术评价 & 复杂性理论	0.000	6.889	6.889
谣言传播 & 说服传播理论	0.000	6.757	6.757

续表

跨学科理论结合增长点（top10）	引用情感系数		引用情感系数增加值
	2020 年	2021 年	
科学数据 &FAIR 原则	0.000	62.138	62.138
科学数据 & 科学数据管理理论	0.000	44.524	44.524
技术创新 &TOE 框架	0.000	14.639	14.639
隐私政策 & 认知心理学	0.000	11.941	11.941
突发事件 & 社会影响理论	0.000	9.262	9.262
影响因素 & 系统动力学	0.000	8.801	8.801
信息搜寻 & 信息觅食理论	0.000	8.733	8.733
技术生命周期 & 系统动力学	0.000	8.000	8.000
满意度 &D and M 框架	0.000	7.594	7.594
情报能力 & 能力理论	0.000	7.372	7.372

表 7 统计了跨学科方法结合点年度增长分布情况。图情领域跨学科方法主要来自计算机科学领域，并且随着计算机科学研究中技术方法的突破与进步，跨学科知识结合点也相应发生着变化。2017 年到 2018 年度，引入的跨学科方法主要来自计算机科学中浅层机器学习方法。2018 年到 2019 年度，主要关注计算机领域中自然语言处理技术和情感分析技术在图情领域术语抽取和评价模型问题中的应用。2019 年到 2020 年度，关注的是机器学习方法在专利和技术创新问题上的运用。2020 年到 2021 年度，随着人工智能研究的火热，图情领域跨学科方法结合点也开始转向深度学习和神经网络。

表 7　跨学科方法结合点年度增长分布表

跨学科方法结合增长点（top10）	引用情感系数		引用情感系数增加值
	2017 年	2018 年	
搜索行为 & 心智模型	0.000	22.606	22.606
人文社会科学 & 一维 PPC 模型	0.000	18.225	18.225
术语抽取 & 机器学习方法	0.000	16.000	16.000
科技文献 & 语义标注	0.000	15.321	15.321
技术创新 & 文本挖掘	1.000	15.177	14.177
影响力 & 朴素贝叶斯	0.000	14.160	14.160
影响力 & 卡尔曼滤波	0.000	14.160	14.160
影响力 & 支持向量机	0.000	14.160	14.160
注意力机制 & 决策模型	0.000	13.019	13.019
注意力机制 & 卷积神经网络	0.000	13.019	13.019
跨学科方法结合增长点（top10）	引用情感系数		引用情感系数增加值
	2018 年	2019 年	
术语抽取 & 依存句法分析	0.000	24.156	24.156
术语抽取 & 情感分析	0.000	18.175	18.175
术语抽取 & 自然语言处理	0.000	18.175	18.175
评价模型 & 情感分析	0.000	17.661	17.661

续表

跨学科方法结合增长点（top10）	引用情感系数		引用情感系数增加值
	2018 年	2019 年	
关键词 & 语义分析	0.000	17.400	17.400
评价模型 & 条件随机场模型	0.000	16.661	16.661
评价模型 & 自然语言处理方法	0.000	16.661	16.661
评价模型 & 情感分类	0.000	16.661	16.661
评价模型 & 认知情感评价模型	0.000	16.661	16.661
评价模型 & 语义分析	0.000	16.661	16.661
跨学科方法结合增长点（top10）	引用情感系数		引用情感系数增加值
	2019 年	2020 年	
技术创新 & 聚类	0.000	33.734	33.734
关键词 & 机器学习	9.542	37.709	28.167
相似度 &LDA	0.000	27.177	27.177
关键词 & 数据挖掘	2.413	29.193	26.780
技术创新 &LDA	0.000	24.402	24.402
技术创新 & 专利挖掘	1.000	25.402	24.402
关键词 & 语义处理	6.062	30.193	24.131
技术创新 & 专利文本挖掘	0.000	22.402	22.402
技术创新 & 词嵌入算法	0.000	22.402	22.402
技术创新 &LDA 主题模型	0.000	22.402	22.402
跨学科方法结合增长点（top10）	引用情感系数		引用情感系数增加值
	2020 年	2021 年	
科学数据 & 系统模型	0.000	44.524	44.524
科学数据 & 管理模型	0.000	44.524	44.524
政策量化 & 评价指标体系	0.000	18.194	18.194
信息资源 & 区块链技术	1.000	18.828	17.828
技术创新 & 相似性	0.000	16.703	16.703
注意力机制 & 神经网络	1.000	17.500	16.500
术语抽取 & 神经网络	0.000	15.989	15.989
大数据分析 & 管理体系	0.000	15.465	15.465
术语抽取 &LSTM-CRF	0.000	15.106	15.106
知识抽取 &BERT	0.000	15.012	15.012

表 8 统计了跨学科工具结合点年度增长分布情况，从中可以看出图情领域使用较多的跨学科工具主要为计算机语料工具、可视化工具和数据分析软件工具。总体来看，随着学科研究热点的转化，图情学科使用的工具会发生快速更新。2017 年到 2018 年度，图情学科研究对象主要为社交媒体，研究中使用的数据工具平台多为网络百科、微信和微博。2018 年到 2019 年度，随着图情学科情感分析方法使用频率的增加，情感词典工具的使用逐渐增多。2019 年到 2020 年度，随着政府数据开放、技术创新等研究方向的火热，政府门户网站、专利数据库以及问卷工具的使用逐渐增多。2020 年到 2021 年度，科学数据研究的兴起使得开放数据平台等使用更为频繁。

表 8　跨学科工具结合点年度增长分布表

跨学科工具结合增长点（top10）	引用情感系数		引用情感系数增加值
	2017 年	2018 年	
标准化 & 电子阅读器	0.000	15.949	15.949
理论模型 & 可视化	0.000	15.264	15.264
学术期刊 & 微信	1.000	16.000	15.000
政务微博 & 微博	7.515	22.039	14.525
政务微信 & 微信	1.000	15.142	14.142
数据政策 & 政策工具	0.000	13.341	13.341
注意力机制 & 推文	0.000	13.019	13.019
影响因素 & 社交网站	6.174	19.149	12.975
多媒体 & 电子阅读器	0.000	12.665	12.665
技术创新 & 创新政策	3.385	15.679	12.294
情报学 & 可视化	0.000	19.917	19.917
跨学科工具结合增长点（top10）	引用情感系数		引用情感系数增加值
	2018 年	2019 年	
评价模型 & 情感词典	0.000	16.661	16.661
评价模型 & 社交网络	1.293	16.661	15.368
评价模型 & 微博	1.746	16.661	14.915
关键词 & 专利地图	0.000	14.213	14.213
比较分析 &GO	0.000	13.768	13.768
比较分析 &Wikidata	0.000	13.768	13.768
比较分析 &DBpedia	0.000	13.768	13.768
比较分析 &YAGO	0.000	13.768	13.768
网络用户 &Twitter	2.000	15.663	13.663
跨学科工具结合增长点（top10）	引用情感系数		引用情感系数增加值
	2019 年	2020 年	
技术创新 & 专利数据	5.452	37.204	31.752
政府数据开放 & 数据开放平台	4.000	34.033	30.033
关键词 & 语义网络	0.000	29.193	29.193
模型构建 & 调查问卷	0.000	24.447	24.447
技术创新 & 语义网络	0.000	22.402	22.402
技术创新 & 可视化	0.000	22.402	22.402
影响因素 & 政策工具	0.000	18.175	18.175
政府数据开放 & 门户网站	0.000	18.151	18.151
隐私关注 & 微信	0.000	16.978	16.978
相似度 & 社交网络	0.000	16.628	16.628
跨学科工具结合增长点（top10）	引用情感系数		引用情感系数增加值
	2020 年	2021 年	
科学数据 & 开放研究数据平台	0.000	44.524	44.524
科学数据 &Google Scholar	0.000	44.524	44.524
科学数据 &RDA	1.000	44.524	43.524
分析框架 & 政策工具	3.770	23.620	19.850
政策量化 & 创新政策	0.000	18.194	18.194
科学数据 &GO	0.000	15.503	15.503
专利价值 & 专利数据	1.000	15.835	14.835
技术创新 & 高端市场	0.000	14.789	14.789
信息检索 &APP	0.000	14.030	14.030
社交媒体 & 可视化	1.263	14.609	13.347

3.3　不同类型跨学科知识结合生长点识别

本文借鉴生命周期理论[31]，通过观察跨学科知识结合点与引用情感系数之间存在的关联性，将跨学科知识结合点的生命周期划分为“形成期”“成长期”和“成熟期”三个阶段。(1) 形成期。目标学科知识尝试与多个跨学科知识点结合，未能形成稳固的结合生长点，引用情感系数在低位值之间来回波动。(2) 成长期。目标学科知识与跨学科知识形成稳固的结合生长点，引用情感系数随之快速增长。(3) 成熟期。跨学科知识逐渐被目标学科知识体系吸收和内化，引用情感系数生长曲线也逐渐萎缩。

本文按跨学科知识结合点的不同类型，绘制了 2017 年到 2021 年引用情感系数增长排名前 10 的年度分布图，具体如图 5、图 6、图 7、图 8 所示，期望发现图情领域潜在的跨学科结合生长点。综合来看，图情领域跨学科知识结合点大部分生命周期较短，缺乏长期稳定的跨学科知识输入结合点。如图 5 所示，近年来关于政府数据、量化政策、健康信息、专利和信息规避等学科概念的引入，为图情领域研究前沿的扩展打开了方向。结合图 6 可以发现，跨学科理论的引入与目标学科领域中研究热点的变化存在很大的关联性。如图情领域 2021 年较多地引入了管理学、心理学以及社会学前沿理论，这多与图情领域 2021 年科学数据、突发事件、隐私政策以及搜索行为研究的快速发展相关。图 7 展示了图情领域跨学科方法的来源主要为计算机科学和管理学，计算机领域神经网络、区块链研究的深入和突破，为图情领域带来了新的研究算法和模型，随着大数据时代的到来，传统管理学方法在数据管理中的运用也成为了图情领域关注的重点。图 8 显示了图情领域跨学科工具的结合点同样受研究主题的影响，存在较大的波动性。

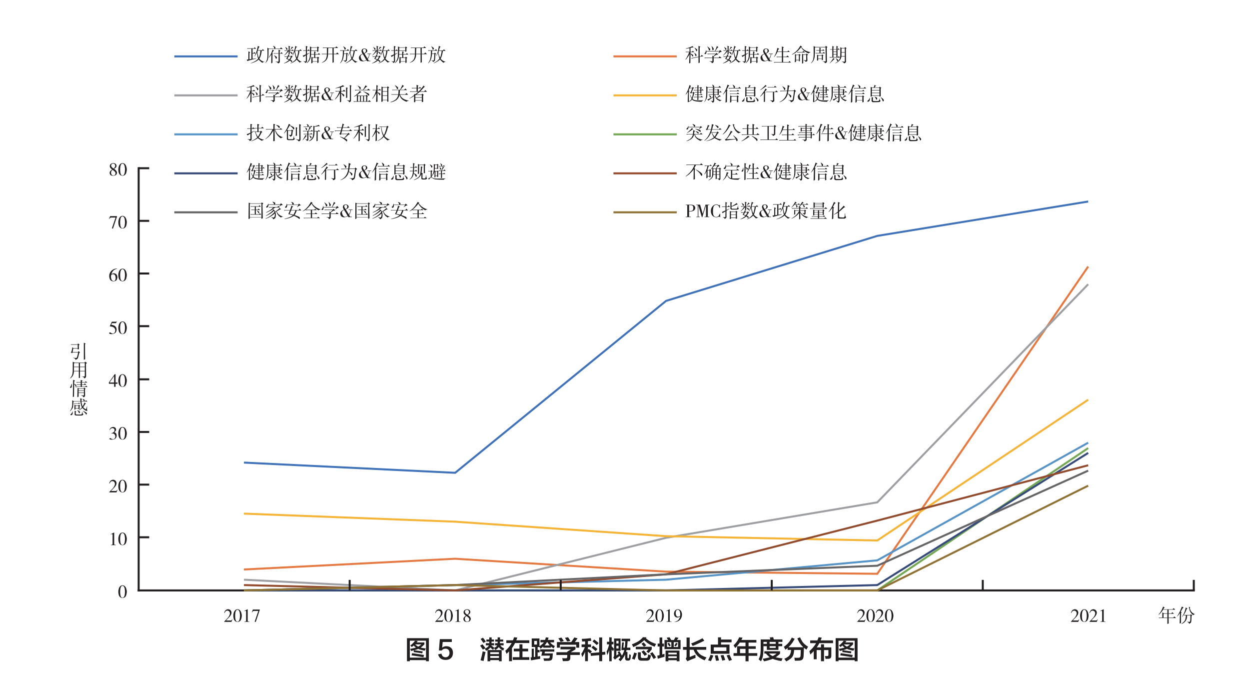


图 5　潜在跨学科概念增长点年度分布图

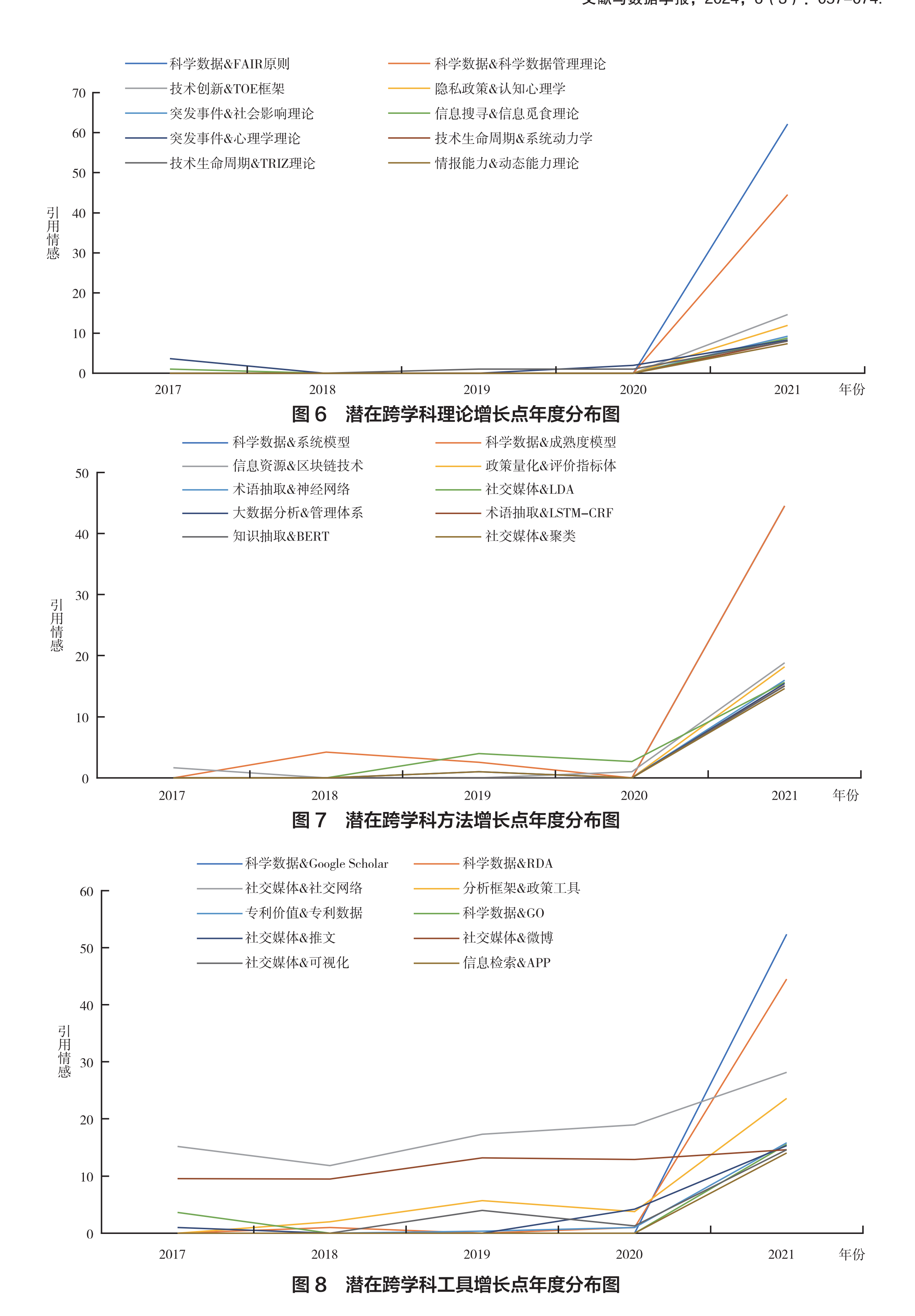


图 6　潜在跨学科理论增长点年度分布图

图 7　潜在跨学科方法增长点年度分布图

图 8　潜在跨学科工具增长点年度分布图

4 总 结

本文通过引用情感分析指标对图情学科中潜在的跨学科知识结合点进行了实证研究，分析了不同类型跨学科知识与图情学科知识结合的生长点状况，发现图情学科研究中跨学科概念、理论和工具结合点的生长受学科领域研究热点的影响很大，而跨学科方法的结合主要依赖于计算机科学中算法、模型的研究进展，并尝试将管理学方法运用在大数据环境下的各类型数据的管理方面。总体来看，图情学科知识与跨学科知识的结合点生命周期较短，缺乏长期稳定的跨学科知识引入点。本文期望通过对我国图情领域跨学科知识结合点的引用情感指标分析，有助广大研究者把握图情学科研究的发展趋势，并且为其他学科潜在跨学科知识生长点的预测与发现提供理论支撑和研究思路。本研究也存在一定局限性，如存在实证研究中文献采集的时间跨度太短、对于引用内容中跨学科知识点和跨学科主题的识别和界定较为粗略等问题，在后期的研究中有待改进。

【参考文献】

[1] 刘仲林. 交叉科学时代的交叉研究 [J]. 科学学研究，1993，11（2）：9–16.

[2] 路甬祥. 学科交叉与交叉科学的意义 [J]. 中国科学院院刊，2005（1）：58–60.

[3] Pierce S J. Boundary crossing in research literatures as a means of interdisciplinary information transfer [J]. Journal of the Association for Information Science and Technology, 1999, 50(3): 271–279.

[4] 刘献君. 学科交叉是建设世界一流学科的重要途径 [J]. 高校教育管理，2020，14（1）：1–7.

[5] 中国大百科全书《生物学》编委会. 中国大百科全书：生物学 3 [M]. 北京：中国大百科全书出版社，1992：257.

[6] 李长玲，高峰，牌艳欣. 试论跨学科潜在知识生长点及其识别方法 [J]. 科学学研究，2021，39（6）：1007–1014.

[7] Porter A L, Cohen A S, David R J, et al.Measuring researcher interdisciplinarity [J]. Scientometrics, 2007, 72(1): 117–147.

[8] 李江. “跨学科性”的概念框架与测度 [J]. 图书情报知识，2014（3）：87–93.

[9] 柯青，朱婷婷. 图书情报学跨学科期刊引用及知识贡献推进效应——基于JCR社会科学版的分析 [J]. 情报资料工作，2017（2）：12–21.

[10] Yegros A, Rafols I, D’Este P, et al. Does interdisciplinary research lead to higher citation impact? The different effect of proximal and distal interdisciplinarity [J]. Plos One, 2015, 10(8): 1–21.

[11] 徐璐，李长玲，荣国阳.期刊的跨学科引用对跨学科知识输出的影响研究——以图书情报领域为例 [J]. 情报杂志，2021，40（7）：182–188.

[12] 周娜，李秀霞，高丹，等. 基于潜在主题的知识组合分析研究——以传播学为例 [J]. 农业图书情报学刊，2018，30（9）：85–90.

[13] 周娜，李秀霞，高丹. 基于LDA主题模型的“作者—内容—方法”多重共现分析——以图书情报学为例 [J]. 情报理论与实践，2019，42（6）：144–148，123.

[14] 李长玲，荣国阳，范晴晴，等. 跨学科引用对知识生长的刺激作用与程度分析——以图书情报学为例 [J]. 情报理论与实践，2022，45（6）：1–8.

［15］李长玲，范晴晴，荣国阳，等. 动能理论视角下跨学科知识生长点成长态势分析——以图书情报领域为例［J］. 情报理论与实践，2023，46（3）：9-15.

［16］荣国阳，李长玲，范晴晴，等. 基于生命周期理论的跨学科知识生长点识别——以引文分析领域为例［J］. 情报理论与实践，2022，45（6）：9-16.

［17］温芳芳，郑诗嘉. 基于关联规则挖掘的多学科知识融合研究——以新冠肺炎研究领域为例［J］. 现代情报，2023，43（3）：148-156.

［18］徐庶睿，卢超，章成志. 术语引用视角下的学科交叉测度——以PLOS ONE上六个学科为例［J］. 情报学报，2017，36（8）：809-820.

［19］章成志，徐庶睿，卢超. 利用引文内容监测多学科交叉现象的方法与实证［J］. 图书情报工作，2016，60（19）：108-115.

［20］Moravcsik M J, Murugesan P. Some results on the function and quality of citations［J］. Social Studies of Science, 1975, 5(1): 86-92.

［21］徐琳宏，丁堃，林原，等. 基于机器学习算法的引文情感自动识别研究——以自然语言处理领域为例［J］. 现代情报，2020，40（1）：35-40，48.

［22］Agarwal A, Rambow O. In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing（EMNLP'10）, October, 2010［C］. Cambridge, MA: Association for Computational Linguistics, 2010.

［23］廖君华，刘自强，白如江，等. 基于引文内容分析的引用情感识别研究［J］. 图书情报工作，2018，62（15）：112-121.

［24］左任衔，唐振华，黄晓，等. 基于情感词典的引文文本情感识别研究［J］. 数字图书馆论坛，2022（2）：10-17.

［25］王芳，陈锋，祝娜，等. 我国情报学理论的来源、应用及学科专属度研究［J］. 情报学报，2016，35（11）：1148-1164.

［26］赵洪，王芳. 理论术语抽取的深度学习模型及自训练算法研究［J］. 情报学报，2018，37（9）：923-938.

［27］Guo H, Cherry C. In Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, May-June, 2015［C］. Denver, Colorado: Association for Computational Linguistics, 2015.

［28］姜霖，张麒麟. 基于引文细粒度情感量化的学术评价研究［J］. 数据分析与知识发现，2020，4（6）：129-138.

［29］苏丽敏，何慧爽. 基于区间数的Spearman秩相关系数的多属性决策方法［J］. 统计与决策，2019，35（6）：51-53.

［30］耿树青，杨建林. 基于引用情感的论文学术影响力评价方法研究［J］. 情报理论与实践，2018，41（12）：93-98.

［31］马费成，望俊成. 信息生命周期研究述评（Ⅰ）——价值视角［J］. 情报学报，2010，29（5）：939-947.

文献与数据学报
第 6 卷　第 3 期　　2024 年 9 月

Identification of Potential Interdisciplinary Knowledge Growth Points Based on Citation Sentiment: Taking Domestic Library and Information Science as Example

Jiang Lin[1]　Zhang Qilin[2]

(1. Business School of Nantong University, Nantong 226019, China;
2. Business College of Southwest University, Chongqing 402460, China)

Abstract: ［**Purpose/Significance**］With the increasing complexity and comprehensiveness of scientific research issues, identifying and studying the interdisciplinary knowledge growth points in current disciplines can help find the current development status and trends of disciplines, and explore the laws of interdisciplinary evolution. ［**Method/Process**］It takes six kinds of academic journals of domestic library and information science as an example and extracts the full text of literature from 2017 to 2021 for empirical analysis. The research process uses the BiLSTM+CRF model to identify specific knowledge combining points in the interdisciplinary citation content, and attempts to evaluate the growth status and trends of interdisciplinary knowledge combining points from the perspective of citation sentiment, thereby providing assistance for the discovery of potential interdisciplinary knowledge growth points. ［**Result/Conclusion**］The results indicate that the growth of interdisciplinary concepts, theories, and tool combining points in the field of library and information science is greatly influenced by the current research hotspots of the field. The combination of interdisciplinary methods mainly depends on the research progress of new algorithm models in computer science, and attempt to apply management methods in the new data environments. Overall, the lifecycle of the combination of disciplinary knowledge and interdisciplinary knowledge in the field of library and information science is relatively short, and there is a lack of long-term and stable interdisciplinary knowledge introduction points.

Keywords: Interdisciplinary; Knowledge growth points; Citation analysis; Citation sentiment

（本文责编：魏　进）